Journal of Informatics and Advanced Computing (JIAC) **E-ISSN : 2722-0346**
Vol. 7 No.1, Mei 2026

Klasterisasi Disparitas Wilayah di Indonesia Menggunakan Indeks Pembangunan Manusia dan Indikator Kemiskinan

Rizki Hesananda[-1], Ari Jaya Ramlah Saputra[-2] , Andiani Abimanyu[-3]
Universitas Siber Indonesia [-1], Bank Rakyat Indonesia [-2] Universitas Pancasila [-2]
email: hessananda@cyber-univ.ac.id [-1] arijayasaputra10 @gmail.com[-2], andiani@univpancasila.ac.id

***Abstract*—** *Pembangunan sosial-ekonomi di Indonesia masih menunjukkan ketimpangan antarwilayah yang signifikan, khususnya dalam hal Indeks Pembangunan Manusia (IPM) dan tingkat kemiskinan. Indikator nasional sering kali tidak mampu menggambarkan ketimpangan pada tingkat kabupaten/kota, sehingga diperlukan pendekatan berbasis data untuk mengidentifikasi pola yang lebih rinci. Penelitian ini bertujuan untuk menerapkan teknik data mining dalam pengelompokan data IPM dan kemiskinan pada 514 kabupaten/kota di Indonesia tahun 2023 guna menghasilkan wawasan yang dapat dimanfaatkan dalam perencanaan pembangunan dan intervensi kebijakan yang lebih tepat sasaran. Penelitian ini menggunakan metodologi CRISP-DM (Cross-Industry Standard Process for Data Mining) dan mengimplementasikan dua algoritma unsupervised learning, yaitu K-Medoids dan DBSCAN. Data diperoleh dari Badan Pusat Statistik (BPS) dan mencakup tiga indikator numerik: IPM, jumlah penduduk miskin, dan persentase penduduk miskin. Data diproses melalui tahapan pembersihan, standarisasi, dan normalisasi agar siap digunakan dalam pemodelan. Algoritma K-Medoids menghasilkan lima klaster dengan nilai Silhouette Coefficient sebesar 0,35 dan Davies-Bouldin Index sebesar 0,86, sementara DBSCAN membentuk tiga label (termasuk 34 data outlier) dengan Silhouette Coefficient sebesar 0,40 dan DBI sebesar 1,34. Masing-masing klaster mencerminkan karakteristik sosial-ekonomi yang unik. Hasil penelitian menunjukkan bahwa teknik clustering mampu memberikan segmentasi yang bermakna terhadap ketimpangan pembangunan. Temuan ini sangat berguna bagi instansi pemerintah, lembaga pembangunan, LSM, dan sektor swasta dalam merumuskan strategi pembangunan yang lebih adil dan berkelanjutan di tingkat daerah..*

Kata Kunci : *Indeks Pembangunan Manusia (IPM), Kemiskinan, Data Mining, Klasterisasi, CRISP-DM*

1. PENDAHULUAN

Pembangunan manusia dan kemiskinan tetap menjadi dua isu yang paling krusial dan saling berkaitan dalam konteks pembangunan global maupun nasional. Meskipun berbagai kemajuan telah dicapai dalam peningkatan pendidikan, kesehatan, dan kesejahteraan ekonomi di banyak wilayah, kesenjangan antara daerah yang mengalami kemajuan dan daerah yang mengalami stagnasi masih terus terjadi, terutama di negara berkembang seperti Indonesia [1], [2]. Disparitas ini menunjukkan pentingnya analisis indikator pada tingkat regional yang melampaui agregasi nasional untuk memastikan bahwa pembangunan berlangsung secara merata dan inklusif bagi seluruh lapisan masyarakat.

Indeks Pembangunan Manusia (IPM) secara luas diakui sebagai indikator komposit yang mencerminkan tingkat kesejahteraan masyarakat melalui dimensi pendidikan, kesehatan, dan standar hidup [3], [4]. Di sisi lain, indikator kemiskinan baik dalam bentuk jumlah penduduk miskin maupun tingkat kemiskinan tetap menjadi ukuran utama dalam mengevaluasi kondisi sosial ekonomi suatu wilayah. Dalam konteks Indonesia, variasi regional pada IPM dan kemiskinan masih sangat signifikan [5]. Wilayah perkotaan seperti Jakarta dan Yogyakarta secara konsisten menunjukkan nilai IPM yang tinggi dan tingkat kemiskinan yang rendah, sementara daerah pedesaan dan wilayah terpencil, khususnya di kawasan Indonesia Timur, menghadapi kondisi yang sebaliknya [6], [7].

Meskipun IPM dan kemiskinan sering dianalisis secara terpisah, hubungan antara kedua indikator tersebut memberikan pemahaman yang lebih komprehensif mengenai dinamika pembangunan wilayah [8]. Penelitian terdahulu menunjukkan bahwa IPM yang rendah sering kali berkorelasi dengan tingkat kemiskinan yang lebih tinggi, namun hubungan tersebut tidak selalu bersifat linear karena dipengaruhi oleh struktur ekonomi lokal, tata kelola pemerintahan, serta faktor budaya [9], [10]. Oleh karena itu, identifikasi pola yang mengintegrasikan ked[11]ua dimensi tersebut menjadi penting dalam perumusan kebijakan publik yang efektif, terutama dalam konteks tata kelola pemerintahan yang terdesentralisasi seperti di Indonesia .

Dengan semakin tersedianya data yang terbuka dan rinci, teknik data mining menawarkan pendekatan yang efektif untuk mengekstraksi pengetahuan dari kumpulan data yang besar dan kompleks [12]. Salah satu teknik utama dalam pembelajaran tanpa supervisi (unsupervised learning) adalah klasterisasi (clustering), yang memungkinkan pengelompokan entitas berdasarkan tingkat kemiripan atribut tertentu [13], [14]. Dalam penelitian ini, klasterisasi diterapkan untuk mengelompokkan kabupaten dan kota di Indonesia berdasarkan indikator IPM dan kemiskinan. Pendekatan ini memungkinkan peneliti dan pembuat kebijakan untuk mengidentifikasi kelompok-kelompok alami (*natural clusters*) yang mungkin

tidak terlihat melalui metode klasifikasi konvensional .

Untuk memandu pelaksanaan proses klasterisasi tersebut, penelitian ini mengadopsi metodologi CRISP-DM (*Cross-Industry Standard Process for Data Mining*). CRISP-DM menyediakan model proses yang terstruktur yang terdiri atas enam tahapan, yaitu pemahaman bisnis (*business understanding*), pemahaman data (*data understanding*), persiapan data (*data preparation*), pemodelan (*modeling*), evaluasi (*evaluation*), dan implementasi (*deployment*) [15]. Dengan mengikuti kerangka kerja ini, penelitian dapat menjaga ketelitian metodologis sekaligus memastikan keselarasan antara analisis teknis dan tujuan penelitian. Selain itu, penggunaan algoritma K-Medoids dan DBSCAN memungkinkan perbandingan antara pendekatan klasterisasi berbasis centroid dan berbasis kepadatan (density-based), yang masing-masing memiliki keunggulan dalam menangani data sosial ekonomi dunia nyata [16], [17].

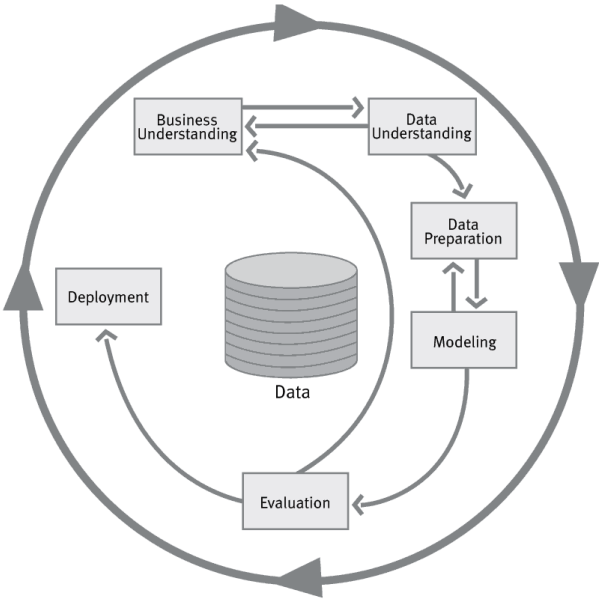
Literatur yang ada telah banyak mengeksplorasi penggunaan teknik klasterisasi dalam studi pembangunan wilayah, namun sebagian besar masih terbatas pada data tingkat provinsi atau hanya menggunakan satu algoritma, seperti K-Means. Penelitian ini memperluas cakupan analisis dengan menggunakan data seluruh kabupaten dan kota di Indonesia serta mengevaluasi kinerja beberapa algoritma menggunakan metrik validitas internal, seperti Silhouette Coefficient dan Davies-Bouldin Index [18]. Penggunaan metrik tersebut bertujuan untuk memastikan bahwa klaster yang dihasilkan tidak hanya valid secara statistik, tetapi juga bermakna dalam mendukung perancangan kebijakan.

Kontribusi penelitian ini terletak pada potensinya dalam menghasilkan peta klasterisasi yang praktis untuk menggambarkan keragaman kondisi pembangunan di Indonesia. Setiap klaster dianalisis tidak hanya berdasarkan karakteristik statistik, tetapi juga ditinjau dari perspektif relevansi kebijakan sehingga menghasilkan rekomendasi yang lebih spesifik sesuai dengan karakteristik sosial ekonomi masing-masing kelompok. Penelitian ini diharapkan dapat mendukung proses perumusan kebijakan berbasis bukti (*evidence-based policymaking*) melalui penyediaan kerangka kerja berbasis data untuk penargetan program pengentasan kemiskinan dan investasi pembangunan manusia pada tingkat regional.

Secara keseluruhan, penelitian ini berupaya mengatasi kesenjangan dalam analisis pembangunan wilayah melalui integrasi indikator multidimensi, penerapan metode data mining yang kuat, serta penggunaan kerangka analisis yang terstruktur. Temuan penelitian diharapkan dapat memperkaya diskursus akademik sekaligus memberikan kontribusi praktis dalam bidang perencanaan wilayah, kebijakan sosial, dan tata kelola pemerintahan berbasis data.

2. METODE PENELITIAN

Penelitian ini menggunakan metodologi CRISP-DM (*Cross-Industry Standard Process for Data Mining*) sebagai kerangka kerja yang terstruktur dan bersifat iteratif untuk melakukan analisis klasterisasi terhadap indikator Indeks Pembangunan Manusia (IPM) dan kemiskinan di Indonesia. Pada Gambar 1, ditampilkan Model CRISP-DM terdiri atas enam fase utama, yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*. Penerapan metodologi ini memastikan adanya keselarasan yang jelas antara tujuan penelitian, proses prapemrosesan data, pemodelan algoritma, serta interpretasi hasil untuk mendukung perencanaan kebijakan pembangunan.



Gambar 1. Metode CRISP-DM

2.1. *Business Understanding* (Pemahaman Bisnis)

Fase awal bertujuan untuk mendefinisikan tujuan klasterisasi tingkat IPM dan kemiskinan pada kabupaten/kota di Indonesia guna mengidentifikasi disparitas wilayah serta menghasilkan informasi yang dapat digunakan dalam perumusan kebijakan. Motivasi penelitian ini berasal dari kompleksitas kondisi sosial ekonomi Indonesia, di mana ketimpangan pembangunan manusia dan kemiskinan masih terjadi antarwilayah. Proses klasterisasi diharapkan dapat membantu pemerintah pusat, pemerintah daerah, organisasi non-pemerintah, maupun sektor swasta dalam merancang program pengentasan kemiskinan yang lebih tepat sasaran dan efektif. Dalam konteks ini, tujuan bisnis diterjemahkan ke dalam identifikasi kelompok wilayah yang homogen berdasarkan indikator IPM dan kemiskinan sehingga dapat mendukung intervensi berbasis bukti (*evidence-based intervention*).

2.2 *Data Understanding* (Pemahaman Data)

Pada fase ini dilakukan pengumpulan dan pemeriksaan dataset yang relevan untuk menilai struktur, kualitas, dan kesesuaiannya terhadap kebutuhan analisis klasterisasi. Penelitian menggunakan data sekunder yang tersedia secara publik dan diperoleh dari portal resmi Badan Pusat Statistik (BPS). Dataset mencakup tiga indikator kuantitatif utama tahun 2023, yaitu nilai IPM, jumlah penduduk miskin, dan persentase penduduk miskin untuk masing-masing dari 514 kabupaten/kota di Indonesia. Selain itu, dilakukan analisis eksplorasi data (*Exploratory Data Analysis/EDA*) untuk memahami distribusi variabel, mendeteksi anomali, serta mengamati hubungan potensial antarindikator. Tahapan ini penting untuk memperoleh pemahaman awal terhadap karakteristik data sebelum memasuki proses pemodelan.

2.3. *Data Preparation* (Persiapan Data)

Untuk memastikan data siap digunakan dalam proses pemodelan, dilakukan beberapa tahapan prapemrosesan data. Atribut yang digunakan dalam penelitian ini meliputi nilai IPM, jumlah penduduk miskin, dan persentase penduduk miskin yang diekstraksi dari dataset mentah. Konsistensi satuan data dijaga dengan mengonversi jumlah penduduk miskin dari satuan ribuan menjadi nilai individu sebenarnya. Selanjutnya dilakukan pemeriksaan data hilang (*missing values*) yang menunjukkan bahwa dataset berada dalam kondisi lengkap tanpa nilai yang hilang. Proses transformasi juga dilakukan untuk mengubah atribut kategorikal menjadi representasi numerik yang dapat diproses oleh algoritma. Selain itu, fitur distandarisasi dan dinormalisasi menggunakan metode *StandardScaler* dan *MinMaxScaler* agar kompatibel dengan algoritma klasterisasi berbasis jarak. Tahapan ini memastikan setiap atribut memiliki kontribusi yang seimbang dalam pembentukan klaster sehingga mengurangi bias akibat perbedaan skala data.

2.4. *Modeling* (Pemodelan)

Penelitian ini menerapkan dua algoritma klasterisasi untuk mengelompokkan wilayah, yaitu K-Medoids dan DBSCAN. Algoritma K-Medoids dipilih karena memiliki ketahanan yang lebih baik terhadap *noise* dan lebih mampu menangani data yang tidak membentuk pola klaster berbentuk bulat (*non-spherical clusters*), sehingga sesuai untuk data sosial ekonomi yang heterogen. Penentuan jumlah klaster optimal (*k*) dilakukan menggunakan metode Elbow. Sementara itu, algoritma DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) digunakan untuk mengeksplorasi kemampuannya dalam mengidentifikasi klaster dengan bentuk dan kepadatan yang beragam tanpa memerlukan jumlah klaster yang telah ditentukan sebelumnya. Penyesuaian parameter (*parameter tuning*) pada DBSCAN dilakukan menggunakan grafik *k-distance* untuk menentukan nilai *eps* (*epsilon*) yang optimal, sedangkan parameter *min_samples* ditetapkan sebesar 5 untuk menjaga keseimbangan sensitivitas terhadap kepadatan data. Penggunaan kedua algoritma tersebut memungkinkan dilakukan perbandingan karakteristik hasil klasterisasi antara pendekatan berbasis centroid (*centroid-based clustering*) dan pendekatan berbasis kepadatan (*density-based clustering*).

2.5. *Evaluation* (Evaluasi)

Model klaster yang dihasilkan dievaluasi menggunakan dua metrik validasi internal, yaitu *Silhouette Coefficient* dan *Davies-Bouldin Index* (DBI). *Silhouette Coefficient* digunakan untuk mengukur seberapa baik suatu objek berada dalam klasternya dibandingkan dengan klaster terdekat lainnya. Nilai yang semakin mendekati 1 menunjukkan bahwa objek memiliki tingkat kesesuaian yang tinggi terhadap klaster tempatnya berada dan memiliki pemisahan yang baik dari klaster lain. Sementara itu, *Davies-Bouldin Index* digunakan untuk mengevaluasi tingkat kemiripan rata-rata antar klaster. Nilai DBI yang lebih rendah menunjukkan kualitas klaster yang lebih baik karena memiliki tingkat pemisahan yang tinggi dan tingkat kekompakan yang lebih baik. Melalui penggunaan kedua metrik tersebut, penelitian ini berupaya memberikan evaluasi yang komprehensif terhadap performa algoritma K-Medoids dan DBSCAN, sekaligus mengidentifikasi algoritma yang paling sesuai untuk data sosial ekonomi Indonesia.

2.6. *Deployment* (Implementasi)

Pada fase implementasi, hasil klasterisasi diterjemahkan ke dalam bentuk informasi yang dapat digunakan oleh pemangku kepentingan dalam proses pengambilan keputusan. Setiap klaster dianalisis berdasarkan karakteristik IPM dan kemiskinan yang dimilikinya untuk menghasilkan pemetaan kondisi pembangunan wilayah di Indonesia. Hasil analisis ini dapat dimanfaatkan sebagai dasar dalam penyusunan kebijakan pembangunan daerah, pengalokasian sumber daya, serta perancangan program pengentasan kemiskinan yang lebih terarah. Selain itu, hasil penelitian juga dapat menjadi referensi bagi pemerintah, akademisi, dan organisasi pembangunan dalam memahami pola disparitas regional serta menentukan prioritas intervensi pembangunan berbasis data.

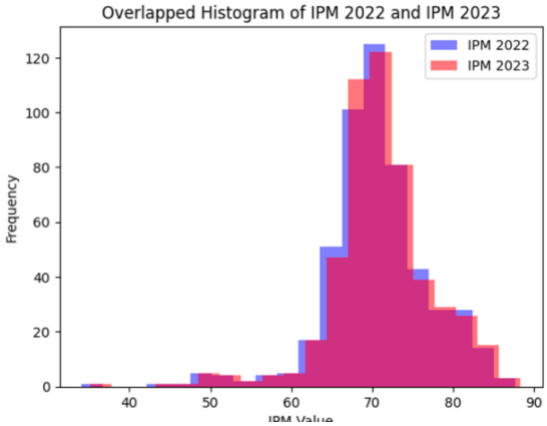
3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil penelitian sesuai dengan enam fase metodologi CRISP-DM. Setiap tahapan berkontribusi terhadap pengembangan model klasterisasi yang mengungkap pola pembangunan manusia dan kemiskinan pada kabupaten/kota di Indonesia.

3.1 *Business Understanding* (Pemahaman Bisnis)

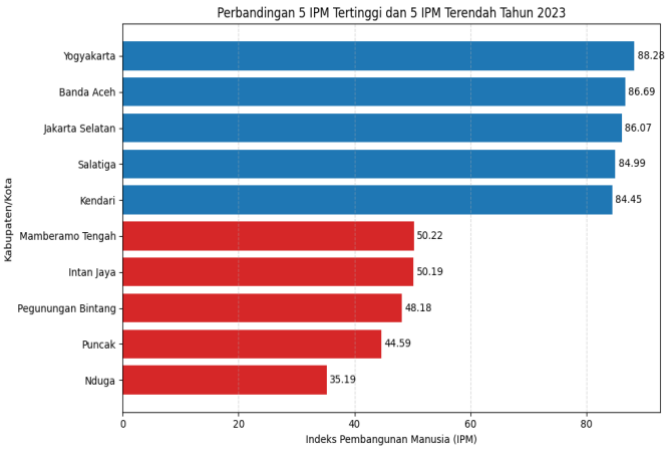
Tahap pertama penelitian ini bertujuan untuk memperoleh pemahaman yang jelas mengenai konteks sosial ekonomi di Indonesia serta pentingnya melakukan klasterisasi data Indeks Pembangunan Manusia (IPM) dan kemiskinan. Indonesia sebagai negara kepulauan yang luas masih menghadapi disparitas pembangunan antarwilayah. Nilai rata-rata IPM dan kemiskinan secara nasional sering kali dapat menyamarkan ketimpangan yang terjadi pada tingkat kabupaten atau kota. Oleh karena itu, tujuan utama penelitian ini adalah menghasilkan kategorisasi wilayah berbasis klasterisasi yang dapat menjadi dasar dalam merumuskan strategi pembangunan yang lebih terarah dan berbeda sesuai dengan karakteristik masing-masing wilayah. Kategorisasi

tersebut penting tidak hanya bagi lembaga nasional seperti Bappenas dan Kementerian Sosial, tetapi juga bagi pemerintah daerah dan organisasi non-pemerintah yang berupaya menyesuaikan program pendidikan, kesehatan, dan kesejahteraan sosial dengan kebutuhan wilayah. Untuk mendukung tujuan tersebut, penelitian ini merumuskan pendekatan berbasis data (*data-driven approach*) untuk mengidentifikasi wilayah yang memiliki karakteristik serupa dalam aspek pembangunan manusia dan kondisi kemiskinan. Salah satu temuan awal yang menjadi dasar analisis diperoleh melalui pengamatan terhadap perkembangan IPM Indonesia dari waktu ke waktu.



Gambar 2. Histogram IPM 2022 dan 2023

Pada Gambar 2, menurut Badan Pusat Statistik (2024), IPM Indonesia pada tahun 2023 mencapai 74,39, meningkat dibandingkan tahun sebelumnya dan mencerminkan peningkatan pada aspek pendidikan, kesehatan, dan standar hidup. Namun, nilai agregat tersebut masih menyembunyikan heterogenitas antarwilayah. Pusat-pusat perkotaan seperti Jakarta dan Yogyakarta terus menunjukkan kinerja yang baik, sedangkan wilayah terpencil, khususnya di kawasan Indonesia Timur, masih mengalami ketertinggalan.



Gambar 3. Perbandingan IPM Kabupaten/Kota

Gambar 3 menunjukkan kontras yang sangat jelas antara lima kabupaten/kota dengan IPM tertinggi dan lima kabupaten/kota dengan IPM terendah pada tahun 2023. Sebagai contoh, Kota Yogyakarta mencatat IPM tertinggi sebesar 88,28, sedangkan Kabupaten Nduga di Papua memiliki IPM yang jauh lebih rendah, yaitu 35,19. Perbedaan yang hampir mencapai dua kali lipat tersebut menunjukkan adanya ketimpangan geografis yang mendalam.

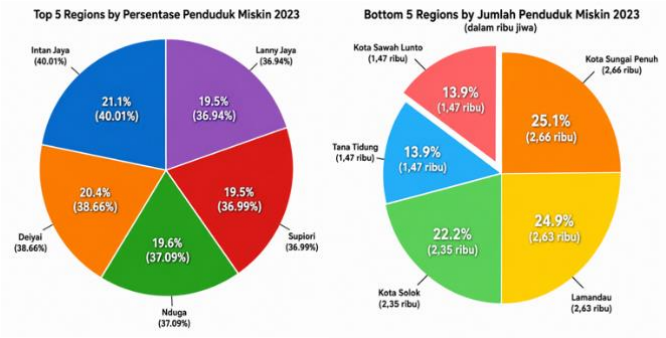
Implikasi dari data tersebut dapat dilihat dari dua aspek. Pertama, kondisi tersebut memperkuat kebutuhan akan klasterisasi karena satu nilai IPM nasional tidak mampu menggambarkan variasi lokal yang signifikan dan memerlukan bentuk intervensi yang disesuaikan dengan karakteristik wilayah. Kedua, kondisi tersebut menjadi titik awal untuk memahami kelompok kabupaten/kota yang memiliki tingkat kinerja serupa sehingga dapat ditangani melalui paket kebijakan yang relatif seragam. Bagi pembuat kebijakan, pendekatan klasterisasi memberikan perspektif yang lebih terperinci dalam mengalokasikan sumber daya, mengevaluasi kinerja pembangunan daerah, serta melakukan intervensi secara lebih efektif pada wilayah yang memiliki tingkat kebutuhan paling tinggi. Dengan demikian, permasalahan bisnis yang diterjemahkan ke dalam tugas data mining adalah mengelompokkan kabupaten/kota di Indonesia berdasarkan indikator IPM dan kemiskinan sebagai dasar dalam penyusunan rekomendasi kebijakan yang spesifik untuk masing-masing karakteristik wilayah.

3.2 *Data Understanding* (Pemahaman Data)

Pada tahap *Data Understanding*, dilakukan pengumpulan dan eksplorasi data untuk memahami karakteristik, kualitas, dan distribusi data yang akan digunakan dalam proses klasterisasi. Dataset penelitian diperoleh dari portal resmi Badan Pusat Statistik (BPS) Indonesia dan mencakup data seluruh 514 kabupaten/kota di Indonesia pada tahun 2023.

Penelitian ini menggunakan tiga variabel utama, yaitu Indeks Pembangunan Manusia (IPM), jumlah penduduk miskin, dan persentase penduduk miskin. Ketiga variabel tersebut dipilih karena mampu merepresentasikan kondisi pembangunan manusia dan tingkat kesejahteraan masyarakat pada tingkat daerah. Hasil pemeriksaan awal menunjukkan bahwa dataset berada dalam kondisi lengkap tanpa nilai yang hilang (*missing values*) sehingga dapat langsung digunakan pada tahap analisis berikutnya.

Selanjutnya, Gambar 4 menunjukkan lima kabupaten dengan persentase penduduk miskin tertinggi pada tahun 2023. Kabupaten Intan Jaya memiliki tingkat kemiskinan tertinggi sebesar 40,01%, diikuti oleh Deiyai, Nduga, Supiori, dan Lanny Jaya yang seluruhnya berada di atas 36%. Temuan ini mengindikasikan bahwa tingkat kemiskinan masih terkonsentrasi pada beberapa wilayah tertentu, khususnya di kawasan Indonesia Timur.



Gambar 4. Lima Kabupaten dengan Persentase Penduduk Miskin Tertinggi Tahun 2023

Hasil eksplorasi data menunjukkan adanya variasi yang cukup besar pada nilai IPM maupun indikator kemiskinan antarwilayah. Kondisi ini mengindikasikan bahwa karakteristik pembangunan daerah di Indonesia bersifat heterogen sehingga diperlukan pendekatan klasterisasi untuk mengelompokkan wilayah-wilayah yang memiliki karakteristik serupa. Dengan demikian, hasil klasterisasi diharapkan dapat memberikan gambaran yang lebih komprehensif mengenai pola pembangunan manusia dan kemiskinan sebagai dasar dalam penyusunan kebijakan yang lebih tepat sasaran.

3.3 Data Preparation (Persiapan Data)

Tahap *Data Preparation* dilakukan untuk menyiapkan data yang akan digunakan dalam proses klasterisasi. Pada tahap ini, dilakukan seleksi atribut yang relevan, pemeriksaan kualitas data, serta transformasi data agar sesuai dengan kebutuhan algoritma K-Medoids dan DBSCAN.

3.3.1 Data Selection

Proses seleksi data dilakukan dengan memilih atribut yang relevan terhadap tujuan penelitian, yaitu mengidentifikasi karakteristik pembangunan manusia dan kemiskinan pada tingkat kabupaten/kota di Indonesia. Dataset yang digunakan berasal dari Badan Pusat Statistik (BPS) dan mencakup 514 kabupaten/kota pada tahun 2023. Atribut yang dipilih terdiri atas Indeks Pembangunan Manusia (IPM), jumlah penduduk miskin, dan persentase penduduk miskin. Ketiga variabel tersebut dipilih karena mampu merepresentasikan kondisi pembangunan manusia dan tingkat kesejahteraan masyarakat secara komprehensif. Hasil seleksi data yang digunakan dalam penelitian ditunjukkan pada Tabel 1.

Tabel 1. Atribut Dataset yang Digunakan dalam Penelitian

No	Atribut	Tipe Data	Keterangan
1	Kabupaten/Kota	Kategorikal	Identitas wilayah
2	IPM 2023	Numerik	Indeks Pembangunan Manusia
3	Jumlah Penduduk Miskin 2023	Numerik	Jumlah penduduk di bawah garis kemiskinan
4	Persentase Penduduk Miskin 2023	Numerik	Persentase penduduk di bawah garis kemiskinan

Berdasarkan hasil seleksi, atribut Kabupaten/Kota digunakan sebagai identitas wilayah, sedangkan atribut IPM, jumlah penduduk miskin, dan persentase penduduk miskin digunakan sebagai variabel dalam proses klasterisasi.

3.3.2 Data Preprocessing dan Transformation

Tahap *preprocessing* dilakukan untuk memastikan kualitas dan konsistensi data sebelum proses pemodelan. Hasil pemeriksaan menunjukkan bahwa dataset tidak memiliki *missing values*, sehingga seluruh data dapat digunakan dalam analisis. Selain itu, variabel jumlah penduduk miskin dikonversi dari satuan ribu jiwa menjadi satuan jiwa untuk menyeragamkan skala pengukuran.

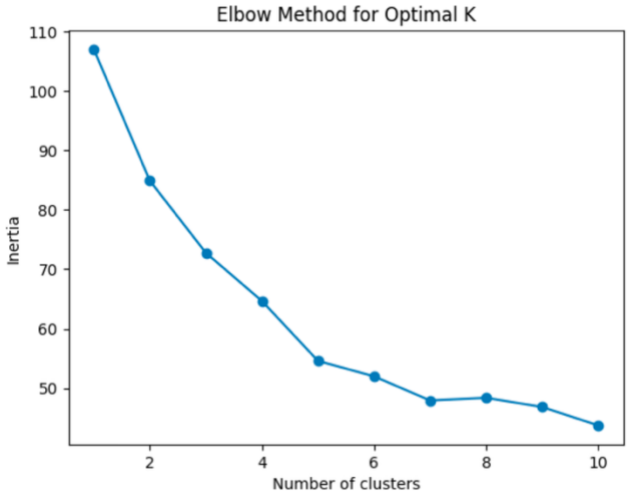
Selanjutnya, data ditransformasikan ke dalam format numerik yang sesuai untuk proses klasterisasi. Standardisasi menggunakan *StandardScaler* dan normalisasi menggunakan *MinMaxScaler* diterapkan untuk mengurangi pengaruh perbedaan rentang nilai antarvariabel. Proses ini memastikan bahwa setiap atribut memberikan kontribusi yang seimbang dalam perhitungan jarak pada algoritma K-Medoids dan DBSCAN.

3.4 *Modeling* (Pemodelan)

Pada tahap pemodelan, algoritma K-Medoids dan DBSCAN diterapkan untuk mengelompokkan 514 kabupaten/kota di Indonesia berdasarkan Indeks Pembangunan Manusia (IPM), jumlah penduduk miskin, dan persentase penduduk miskin tahun 2023. Seluruh proses pemodelan dilakukan menggunakan Python pada Google Colaboratory. Parameter optimal masing-masing algoritma ditentukan terlebih dahulu untuk memperoleh struktur cluster yang representatif.

3.4.1 K-Medoids

K-Medoids digunakan untuk membentuk kelompok wilayah berdasarkan kemiripan karakteristik IPM dan kemiskinan. Jumlah cluster optimal ditentukan menggunakan pendekatan Elbow Method dengan mengevaluasi perubahan nilai inertia pada beberapa kandidat jumlah cluster yang ditunjukan pada Gambar 5. Hasil visualisasi menunjukkan bahwa penurunan inertia mulai melandai pada k = 5, sehingga lima cluster digunakan pada proses pemodelan.



Gambar 5. Penentuan jumlah cluster optimal menggunakan Elbow Method pada K-Medoids.

Berdasarkan nilai tersebut, K-Medoids menghasilkan lima kelompok dengan distribusi data sebagaimana ditunjukkan pada Tabel 2

Tabel 2. Distribusi Kabupaten/Kota pada Cluster K-Medoids

Cluster	Jumlah Kabupaten/Kota	Persentase (%)
Cluster 0	144	28,02
Cluster 1	86	16,73
Cluster 2	80	15,56
Cluster 3	149	28,99
Cluster 4	55	10,70
Total	514	100,00

Hasil tersebut menunjukkan bahwa Cluster 3 memiliki jumlah anggota terbesar, yaitu 149 kabupaten/kota (28,99%), diikuti Cluster 0 sebanyak 144 wilayah (28,02%). Sementara itu, Cluster 4 merupakan kelompok dengan jumlah anggota paling sedikit, yaitu 55 wilayah (10,70%). Perbedaan ukuran cluster menunjukkan adanya variasi karakteristik pembangunan dan kemiskinan antardaerah.

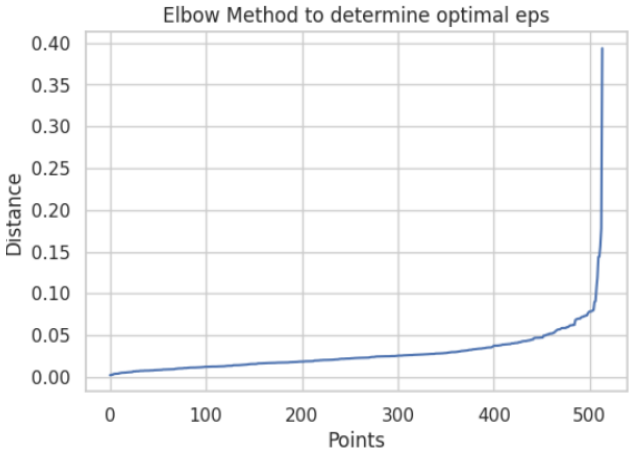
Berbeda dengan DBSCAN yang mampu mengidentifikasi data *noise/outlier* secara eksplisit melalui label -1, algoritma K-Medoids mengharuskan seluruh objek data menjadi anggota salah satu cluster. Oleh karena itu, seluruh 514 kabupaten/kota, termasuk wilayah yang kemudian teridentifikasi sebagai *outlier* oleh DBSCAN, tetap dilibatkan dalam proses pembentukan lima cluster pada K-Medoids.

Salah satu keunggulan K-Medoids dibandingkan K-Means adalah penggunaan objek aktual (*medoid*) sebagai pusat cluster sehingga lebih tahan terhadap pengaruh *outlier*. Meskipun demikian, keberadaan wilayah dengan karakteristik ekstrem, seperti daerah dengan IPM sangat rendah atau tingkat kemiskinan sangat tinggi, tetap dapat memengaruhi komposisi cluster dan meningkatkan variasi internal cluster tertentu.

Dalam penelitian ini, data *outlier* tidak dihapus sebelum proses clustering karena tujuan penelitian adalah memetakan seluruh kondisi kabupaten/kota di Indonesia secara komprehensif. Oleh karena itu, keberadaan wilayah ekstrem dianggap sebagai bagian dari fenomena pembangunan yang perlu dianalisis dan tidak dikeluarkan dari dataset.

3.4.2 DBSCAN

DBSCAN diterapkan sebagai pendekatan berbasis kepadatan untuk mengidentifikasi kelompok wilayah sekaligus mendeteksi data yang memiliki karakteristik berbeda dari kelompok utama. Penentuan nilai *epsilon* dilakukan menggunakan analisis k-distance, dengan *min_samples* ditetapkan sebesar 5 . pada Gambar 6 ditampilkan hasil visualisasi menunjukkan titik perubahan yang digunakan sebagai dasar pemilihan eps = 0,080. Perlu diperhatikan bahwa nilai *epsilon (eps)* sebesar 0,080 diperoleh setelah seluruh variabel numerik melalui proses standarisasi dan normalisasi menggunakan MinMaxScaler sehingga rentang nilai setiap fitur berada pada skala 0 sampai 1. Oleh karena itu, nilai *eps* yang diperoleh bersifat spesifik terhadap dataset hasil normalisasi pada penelitian ini dan tidak dapat langsung diterapkan pada dataset dengan skala asli tanpa proses normalisasi yang sama.



Gambar 6. Penentuan nilai eps pada algoritma DBSCAN.

Dengan parameter eps = 0,080 dan min_samples = 5, DBSCAN menghasilkan satu cluster utama, satu cluster kecil, serta sejumlah data yang dikategorikan sebagai *noise*. Distribusi hasil clustering ditampilkan pada Tabel 3.

Tabel 3. Distribusi Kabupaten/Kota pada Cluster DBSCAN

Label	Jumlah Kabupaten/Kota	Persentase (%)
-1 (Noise)	34	6,61
Cluster 0	475	92,41
Cluster 1	5	0,97
Total	514	100,00

Sebagian besar wilayah, yaitu 475 kabupaten/kota (92,41%), berada pada Cluster 0. Sebanyak 34 wilayah (6,61%) dikategorikan sebagai *noise*, sedangkan Cluster 1 hanya terdiri atas 5 wilayah (0,97%). Temuan ini menunjukkan bahwa sebagian besar kabupaten/kota memiliki karakteristik yang relatif berdekatan, sementara sejumlah wilayah memiliki karakteristik yang cukup berbeda sehingga tidak termasuk dalam cluster utama.

Secara keseluruhan, kedua algoritma menghasilkan struktur pengelompokan yang berbeda. K-Medoids membentuk lima kelompok yang lebih terdistribusi, sedangkan DBSCAN menghasilkan satu kelompok dominan dan mampu mengidentifikasi sejumlah wilayah sebagai noise. Perbedaan struktur tersebut menjadi dasar untuk tahap evaluasi guna menentukan pendekatan yang memberikan kualitas clustering paling sesuai dengan karakteristik data.

3.5 *Evaluation* (Evaluasi)

Evaluasi dilakukan untuk membandingkan kualitas hasil clustering yang dihasilkan oleh algoritma K-Medoids dan DBSCAN. Dua metrik digunakan dalam penelitian ini, yaitu Silhouette Coefficient dan Davies–Bouldin Index (DBI). Silhouette Coefficient digunakan untuk mengukur tingkat kedekatan suatu objek dengan cluster-nya dibandingkan dengan cluster lainnya, sedangkan DBI digunakan untuk mengevaluasi tingkat kekompakan (*compactness*) dan pemisahan (*separation*) antarcluster. Nilai Silhouette yang lebih tinggi dan

nilai DBI yang lebih rendah menunjukkan kualitas clustering yang lebih baik. Hasil evaluasi kedua algoritma ditunjukkan pada Tabel 4.

Tabel 4. Perbandingan Hasil Evaluasi Clustering

Algoritma	Parameter	Silhouette Coefficient	Davies–Bouldin Index (DBI)
K-Medoids	k = 5	0,35	0,86
DBSCAN	eps = 0.080, min_samples = 5	0,40	1,34

Hasil menunjukkan bahwa DBSCAN memperoleh nilai Silhouette Coefficient sebesar 0,40, lebih tinggi dibandingkan K-Medoids sebesar 0,35. Nilai tersebut menunjukkan bahwa pemisahan objek berdasarkan kedekatan internal cluster pada DBSCAN sedikit lebih baik. Namun, DBSCAN menghasilkan nilai DBI sebesar 1,34, sedangkan K-Medoids menghasilkan DBI sebesar 0,86. Karena nilai DBI yang lebih rendah menunjukkan cluster yang lebih kompak dan memiliki pemisahan yang lebih baik, K-Medoids memberikan hasil yang lebih baik berdasarkan metrik tersebut.

Perbedaan kedua metrik menunjukkan bahwa tidak terdapat satu algoritma yang unggul pada seluruh aspek evaluasi. DBSCAN memiliki keunggulan pada Silhouette Coefficient, sedangkan K-Medoids memiliki nilai DBI yang lebih rendah. Dengan mempertimbangkan kedua metrik secara simultan, K-Medoids menunjukkan performa yang lebih seimbang dalam membentuk struktur cluster pada data kabupaten/kota Indonesia. Oleh karena itu, hasil clustering K-Medoids digunakan sebagai dasar utama untuk analisis karakteristik masing-masing kelompok pada tahap berikutnya.

Setelah proses clustering dan evaluasi dilakukan, analisis karakteristik cluster dilakukan untuk menginterpretasikan profil masing-masing kelompok kabupaten/kota berdasarkan nilai rata-rata Indeks Pembangunan Manusia (IPM), jumlah penduduk miskin, dan persentase penduduk miskin. Analisis difokuskan pada hasil K-Medoids karena algoritma tersebut menghasilkan struktur cluster yang lebih seimbang berdasarkan kombinasi nilai Silhouette Coefficient dan Davies–Bouldin Index.

3.5.1 Karakteristik Cluster K-Medoids

K-Medoids menghasilkan lima cluster dengan karakteristik yang berbeda. Ringkasan karakteristik masing-masing cluster disajikan pada Tabel 5.

Tabel 5. Karakteristik Cluster K-Medoids

Cluster	n	Rata-rata IPM	Rata-rata Penduduk Miskin	Rata-rata Kemiskinan (%)	Karakteristik
0	144	69,811	34.616	13,712	IPM relatif rendah, kemiskinan moderat
1	86	80,293	34.514	5,696	IPM tinggi, kemiskinan rendah
2	80	71,589	155.563	11,458	IPM relatif baik, jumlah penduduk miskin tinggi
3	149	71,121	23.198	6,632	IPM menengah, kemiskinan relatif rendah
4	55	60,910	37.173	28,027	IPM rendah, kemiskinan tinggi

Berdasarkan Tabel 5, Cluster 1 memiliki karakteristik pembangunan paling baik. Cluster ini memiliki rata-rata IPM tertinggi, yaitu 80,293, dengan rata-rata persentase penduduk miskin terendah sebesar 5,696%. Kondisi tersebut menunjukkan bahwa wilayah dalam cluster ini secara umum memiliki capaian pembangunan manusia yang relatif tinggi dan tingkat kemiskinan yang lebih rendah dibandingkan cluster lainnya.

Sebaliknya, Cluster 4 menunjukkan kondisi yang paling tertinggal. Cluster ini memiliki rata-rata IPM terendah sebesar 60,910 dan rata-rata persentase penduduk miskin tertinggi sebesar 28,027%. Meskipun hanya mencakup 55 kabupaten/kota, karakteristik tersebut menunjukkan adanya kelompok wilayah yang menghadapi tantangan pembangunan manusia dan kemiskinan secara bersamaan.

Cluster 0 terdiri atas 144 kabupaten/kota dengan rata-rata IPM sebesar 69,811 dan persentase kemiskinan sebesar 13,712%. Karakteristik tersebut menunjukkan kelompok wilayah dengan tingkat pembangunan manusia relatif menengah dan tingkat kemiskinan yang masih cukup tinggi. Sementara itu, Cluster 3, yang merupakan cluster terbesar dengan 149 kabupaten/kota, memiliki rata-rata IPM sebesar 71,121 dan persentase kemiskinan sebesar 6,632%. Dibandingkan Cluster 0, kelompok ini menunjukkan kondisi kemiskinan yang relatif lebih rendah dengan tingkat pembangunan manusia yang sedikit lebih tinggi.

Cluster 2 memiliki karakteristik yang berbeda dari cluster lainnya. Meskipun rata-rata IPM mencapai 71,589, cluster ini memiliki rata-rata jumlah penduduk miskin paling tinggi, yaitu 155.563 jiwa, dengan rata-rata persentase kemiskinan sebesar 11,458%. Hal ini menunjukkan bahwa jumlah penduduk miskin secara absolut tidak selalu berbanding lurus dengan persentase kemiskinan. Suatu wilayah dapat memiliki persentase kemiskinan yang relatif moderat tetapi tetap memiliki jumlah penduduk miskin yang besar karena ukuran populasinya.

Temuan tersebut memperlihatkan bahwa pengelompokan berdasarkan IPM dan indikator kemiskinan mampu mengidentifikasi profil pembangunan wilayah yang berbeda, mulai dari wilayah dengan pembangunan tinggi dan kemiskinan rendah hingga wilayah dengan IPM rendah dan tingkat kemiskinan tinggi. Selain itu, hasil clustering menunjukkan pentingnya membedakan antara persentase kemiskinan dan jumlah penduduk miskin dalam memahami permasalahan sosial-ekonomi suatu wilayah.

3.5.2 Karakteristik Cluster DBSCAN

Sebagai pembanding, DBSCAN menghasilkan struktur yang berbeda dengan K-Medoids. DBSCAN membentuk satu cluster utama, satu cluster kecil, dan 34 data yang dikategorikan sebagai *noise*. Karakteristik masing-masing kelompok disajikan pada Tabel 6.

Tabel 6. Karakteristik Cluster DBSCAN

Label	n	Rata-rata IPM	Rata-rata Penduduk Miskin	Rata-rata Kemiskinan (%)	Karakteristik
-1 (Noise)	34	62,438	105.186	23,488	Wilayah dengan karakteristik relatif berbeda
0	475	71,922	44.390	10,657	Kelompok mayoritas dengan karakteristik umum
1	5	69,282	247.508	10,030	Jumlah penduduk miskin absolut tinggi

Cluster 0 merupakan kelompok dominan yang mencakup 475 kabupaten/kota (92,41%), dengan rata-rata IPM sebesar 71,922 dan rata-rata persentase kemiskinan sebesar 10,657%. Sementara itu, Cluster 1 hanya terdiri atas lima wilayah, tetapi memiliki rata-rata jumlah penduduk miskin sebesar 247.508 jiwa, yang merupakan nilai tertinggi di antara kelompok DBSCAN.

Sebanyak 34 wilayah dikategorikan sebagai noise. Kelompok ini memiliki rata-rata IPM sebesar 62,438, jumlah penduduk miskin sebesar 105.186 jiwa, dan persentase kemiskinan sebesar 23,488%. Nilai tersebut menunjukkan bahwa wilayah yang dikategorikan sebagai noise memiliki karakteristik yang relatif berbeda dari pola mayoritas data, sehingga tidak dapat ditempatkan secara langsung ke dalam cluster utama berdasarkan kepadatan data.

Hasil DBSCAN memperlihatkan keunggulan algoritma berbasis kepadatan dalam mengidentifikasi observasi yang memiliki karakteristik berbeda, sedangkan K-Medoids memberikan pembagian wilayah yang lebih terstruktur ke dalam lima kelompok. Dengan demikian, kedua algoritma memberikan perspektif yang berbeda terhadap pola pembangunan dan kemiskinan di Indonesia.

3.6.1 Deployment (Implementasi)

Hasil implementasi karakteristik menunjukkan bahwa kondisi pembangunan manusia dan kemiskinan di Indonesia bersifat heterogen. Cluster dengan IPM tinggi tidak selalu memiliki jumlah penduduk miskin absolut yang rendah, sebagaimana terlihat pada Cluster 2 K-Medoids. Sebaliknya, wilayah dengan IPM rendah cenderung menunjukkan persentase kemiskinan yang lebih tinggi, sebagaimana ditunjukkan oleh Cluster 4.

Temuan ini memiliki implikasi penting bagi perumusan kebijakan berbasis karakteristik wilayah. Intervensi untuk wilayah seperti Cluster 1 dapat diarahkan pada pemeliharaan kualitas pembangunan dan penguatan produktivitas, sedangkan wilayah seperti Cluster 4 memerlukan perhatian lebih besar terhadap peningkatan pembangunan manusia dan pengurangan kemiskinan. Untuk Cluster 2, pendekatan kebijakan perlu mempertimbangkan skala populasi karena tingginya jumlah penduduk miskin secara absolut dapat menghasilkan kebutuhan intervensi yang berbeda dibandingkan wilayah dengan persentase kemiskinan tinggi tetapi jumlah penduduk yang lebih kecil.

Dengan demikian, hasil clustering tidak hanya berfungsi sebagai pengelompokan statistik, tetapi juga dapat digunakan sebagai pendukung pengambilan keputusan berbasis karakteristik wilayah. Informasi tersebut dapat membantu pemerintah dan pemangku kepentingan dalam menyusun prioritas program pembangunan yang lebih terarah sesuai dengan kondisi masing-masing kelompok wilayah.

4. SIMPULAN

Penelitian ini berhasil menerapkan metodologi CRISP-DM untuk melakukan analisis clustering terhadap Indeks Pembangunan Manusia (IPM) dan indikator kemiskinan pada 514 kabupaten/kota di Indonesia menggunakan algoritma K-Medoids dan DBSCAN. Penelitian ini bertujuan untuk mengungkap pola-pola tersembunyi dalam disparitas pembangunan wilayah sehingga dapat mendukung perumusan kebijakan publik yang lebih tepat sasaran dan berbasis data. Melalui tahapan *business understanding*, *data understanding*, *data preparation*, dan *modeling*, model clustering yang dibangun berhasil menunjukkan adanya heterogenitas yang signifikan dalam kondisi sosial-ekonomi antarwilayah di Indonesia. Hasil clustering menggunakan K-Medoids menghasilkan lima cluster, sedangkan DBSCAN menghasilkan tiga label cluster (termasuk outlier/noise), yang masing-masing merepresentasikan kombinasi karakteristik IPM dan kemiskinan yang berbeda. Pengelompokan tersebut dapat menjadi dasar strategis dalam perencanaan pembangunan daerah dan penentuan prioritas alokasi sumber daya.

Penelitian ini menggunakan seluruh data kabupaten/kota tanpa menghapus data ekstrem (*outlier*) sebelum proses clustering. Pendekatan ini dipilih untuk mempertahankan representasi kondisi pembangunan nasional secara utuh. Namun demikian, penelitian selanjutnya dapat melakukan perbandingan antara hasil clustering sebelum dan sesudah penghapusan *outlier* untuk mengevaluasi pengaruh data ekstrem terhadap stabilitas cluster yang dihasilkan oleh algoritma K-Medoids maupun DBSCAN.

Dari sisi kinerja algoritma, hasil evaluasi menunjukkan bahwa kedua model memberikan perspektif yang berbeda namun saling melengkapi. Algoritma K-Medoids menghasilkan nilai Silhouette Coefficient sebesar 0,35 dan Davies-Bouldin Index (DBI) sebesar 0,86, yang mengindikasikan tingkat kekompakan cluster yang cukup baik serta pemisahan antarcluster yang relatif jelas. Sementara itu, DBSCAN memperoleh nilai Silhouette Coefficient sebesar 0,40, yang menunjukkan kohesi cluster yang sedikit lebih baik, namun menghasilkan DBI sebesar 1,34, yang mengindikasikan pemisahan cluster yang kurang optimal dibandingkan K-Medoids. Selain itu, DBSCAN mengidentifikasi 34 wilayah sebagai outlier, yang menunjukkan adanya daerah dengan

karakteristik sosial-ekonomi yang ekstrem atau unik sehingga memerlukan perhatian kebijakan yang lebih spesifik.

Temuan penelitian ini memiliki implikasi penting dalam konteks perencanaan pembangunan dan pengambilan keputusan di sektor publik. Bagi pemerintah, khususnya Bappenas dan pemerintah daerah, hasil clustering memberikan dasar yang lebih kuat untuk merancang intervensi pembangunan yang terdiferensiasi sesuai karakteristik wilayah, dibandingkan menerapkan kebijakan yang seragam secara nasional. Sebagai contoh, wilayah dengan IPM tinggi tetapi jumlah penduduk miskin yang besar memerlukan strategi pengentasan kemiskinan perkotaan, sedangkan wilayah dengan IPM rendah dan tingkat kemiskinan tinggi membutuhkan prioritas pada pembangunan infrastruktur dasar, peningkatan akses pendidikan dan kesehatan, serta program perlindungan sosial. Di sektor swasta, hasil clustering dapat dimanfaatkan untuk mengidentifikasi wilayah yang memiliki potensi dampak pembangunan terbesar bagi investasi maupun program tanggung jawab sosial perusahaan (*Corporate Social Responsibility*). Selain itu, lembaga swadaya masyarakat dan organisasi donor dapat menggunakan informasi ini untuk menargetkan program bantuan secara lebih tepat sasaran sehingga meningkatkan efektivitas dan efisiensi intervensi yang dilakukan.

Secara keseluruhan, penelitian ini menegaskan pentingnya penerapan teknik *data mining* dan clustering dalam mendukung pembangunan yang lebih inklusif dan berkeadilan. Dengan mengubah data sosial-ekonomi menjadi informasi yang dapat ditindaklanjuti (*actionable insights*), penelitian ini tidak hanya memberikan kontribusi terhadap pengembangan ilmu pengetahuan, tetapi juga mendukung proses pengambilan keputusan di dunia nyata. Penelitian selanjutnya dapat memperluas cakupan analisis dengan menambahkan indikator lain, seperti tingkat pengangguran, akses terhadap layanan publik, atau risiko lingkungan, serta menerapkan analisis clustering secara longitudinal untuk mengidentifikasi dinamika perkembangan IPM dan kemiskinan antarwilayah dari waktu ke waktu di Indonesia.

DAFTAR PUSTAKA

[1] L. Hasibuan, S. R.-J. P. Pendidikan, and undefined 2020, "Analisis Determinan Indeks Pembangunan Manusia (Ipm) Di Indonesia," *jurnal-lp2m.umnaw.ac.id*, Accessed: Sep. 07, 2026. [Online]. Available: https://jurnal-lp2m.umnaw.ac.id/index.php/JP2SH/article/view/470

[2] N. Putri, S. M.-E. J. Ekonomi, and undefined 2022, "pengaruh indeks pengangguran, indeks pelayanan kesehatan dan indeks pendidikan terhadap Indeks Pembangunan Manusia (IPM) di Kabupaten Bojonegoro," *equity.ubb.ac.id*, doi: 10.33019/equity.v10i1.83.

[3] N. Putri, S. M.-E. J. Ekonomi, and undefined 2022, "pengaruh indeks pengangguran, indeks pelayanan kesehatan dan indeks pendidikan terhadap Indeks Pembangunan Manusia (IPM) di Kabupaten Bojonegoro," *equity.ubb.ac.id*, doi: 10.33019/equity.v10i1.83.

[4] J. Aqilah, R. ∗ Jurusan, E. Pembangunan, F. Ekonomi, U. Negeri, and S. Permalink, "Determinan Indeks Pembangunan Manusia (IPM) Provinsi Nusa Tenggara Timur (NTT)," *pdfs.semanticscholar.org*, vol. 4, no. 1, pp. 1080–1092, 2021, doi: 10.15294/efficient.v4i1.41076.

[5] P. Pengeluaran *et al.*, "Pengaruh pengeluaran bidang pendidikan, kesehatan dan infrastruktur jalan terhadap indeks pembangunan manusia (IPM) di Provinsi Riau," *journal.unilak.ac.id*, vol. 15, no. 1, pp. 1–11, 2022, Accessed: Sep. 03, 2026. [Online]. Available: http://journal.unilak.ac.id/index.php/nia/article/view/7380

[6] S. Najah *et al.*, "Pengaruh Investasi dan indeks pembangunan manusia (IPM) terhadap pertumbuhan ekonomi di provinsi Jawa Timur," *journal.lppmpelitabangsa.id*, vol. 10, no. 01, 2025, doi: 10.37366/jespb.v10i01.2405.

[7] P. Jumlah Penduduk, D. Kemiskinan, E. Khristina Kiha, S. Seran, and H. Trifonia Lau, "Pengaruh jumlah penduduk, pengangguran, dan kemiskinan terhadap indeks pembangunan manusia (ipm) di kabupaten belu," *jurnalintelektiva.com*, Accessed: Sep. 07, 2026. [Online]. Available: https://jurnalintelektiva.com/index.php/jurnal/article/view/426

[8] A. Prameswari *et al.*, "Analisis pengaruh kemiskinan, Indeks Pembangunan Manusia (IPM) dan tenaga kerja terhadap pertumbuhan ekonomi di Jawa Timur," *journal.stiem.ac.id*, vol. 7, no. 2, pp. 168–179, 2021, Accessed: Sep. 07, 2026. [Online]. Available: https://journal.stiem.ac.id/index.php/jurep/article/view/909

[9] J. Penduduk, I. Khairunnisa, F. Yusnita, I. Wina Suryani, M. Panorama, and F. Ekonomi dan Bisnis Islam UIN Raden Fatah Palembang, "Jumlah Penduduk, Pengangguran, Kemiskinan Terhadap Indeks Pembangunan Manusia (Ipm) Sumatera Selatan Tahun 2018-2022," *journal.stiemb.ac.id*, vol. 7, no. 3, p. 2023, Accessed: Sep. 03, 2026. [Online]. Available: https://journal.stiemb.ac.id/index.php/mea/article/view/3557

[10] B. Maulana, M. F.-J. Ekonomi, undefined Bisnis, and undefined 2022, "Pengaruh Indeks Pembangunan Manusia (IPM) Terhadap Pertumbuhan Ekonomi di Provinsi Banten Tahun 2019-2021," *journal.unimar-amni.ac.id*, Accessed: Sep. 03, 2026. [Online]. Available: https://journal.unimar-amni.ac.id/index.php/EBISMEN/article/view/81

[11] M. S.-E. R. J. I. E. Dan and undefined 2022, "Analisis Pengaruh Tingkat Kemiskinan Terhadap Indeks Pembangunan Manusia (IPM) Dengan Model Regresi Linier (Studi Kasus Di Kabupaten Bengkulu Utara," *jurnal.unived.ac.id*, Accessed: Sep. 03, 2026. [Online]. Available: https://jurnal.unived.ac.id/index.php/er/article/view/2647

[12] M. A. Rokhib, "Penggunaan Data Mining Untuk Pengelolaan Data Dalam Bidang Pendidikan," pp. 1–4, 2016.

[13] F. Hidayat, R. Putra, … M. A.-I. J. of, and undefined 2023, "Implementasi clustering k-medoids dalam pengelompokan kabupaten di provinsi aceh berdasarkan faktor yang mempengaruhi kemiskinan," *pdfs.semanticscholar.org*, vol. 5, no. 2, pp. 121–130, 2022, doi: 10.13057/ijas.v5i2.55080.

[14] R. Hesananda, H. L. H. S. Warnars, and Sianipar, "Supervised Classification Karakter Morfologi Tanaman Keladi Tikus (Typhonium Flagelliforme) Menggunakan Database Management System," *Jurnal Sistem Komputer*, vol. 7, no. 2, pp. 50–58, 2017, [Online]. Available: https://core.ac.uk/download/pdf/236215548.pdf

[15] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," *Procedia Comput. Sci.*, vol. 181, pp. 526–534, 2021.

[16] M. Fuchs and W. Höpken, "Clustering: Hierarchical, k-Means, DBSCAN," *Tourism on the Verge*, vol. Part F1051, pp. 129–149, 2022, doi: 10.1007/978-3-030-88389-8_8.

[17] E. Herman, K. Zsido, V. F.-A. Sciences, and undefined 2022, "Cluster analysis with k-mean versus k-medoid in financial performance evaluation," *mdpi.com*, Accessed: Sep. 03, 2026. [Online]. Available: https://www.mdpi.com/2076-3417/12/16/7985

[18] S. Huang, Z. Kang, Z. Xu, and Q. Liu, "Robust deep k-means: An effective and simple method for data clustering," *Pattern Recognit.*, vol. 117, p. 107996, 2021.